

Publication Date: 10 August 2024

Archs Sci. (2024) Volume 74, Issue S1 Pages 160-169, Paper ID 2024s122.
<https://doi.org/10.62227/as/74s122>

Research On The Innovation Path Of College Basketball Teaching Based On Big Data Analysis

Bo Zhao¹ and Dongjin Xian^{2,*}

¹Changzhou University, Changzhou, Jiangsu, 213016, China.

²School of Physical Education, Changzhou University, Changzhou, Jiangsu, 213164, China.

Corresponding authors: (e-mail: 13685212719@139.com).

Abstract Basketball, as a confrontational sport, is beneficial to improve the participants' fitness level, flexibility, and teamwork awareness. Based on the educational big data analysis process, this paper collects the current college basketball teaching data from six dimensions, and optimizes the selection of the initial class cluster centroids of the K-means clustering algorithm, and, using the improved DT-k means clustering algorithm, carries out a cluster analysis of the college basketball teaching situation, and explores the current college basketball teaching level. On this basis, this paper designs a new model of college basketball teaching based on SPOC, and at the same time applies it to a college, and proves the teaching effect of the model through a controlled experiment. The experimental results show that the experimental group and the control group of free throw line one minute shooting test and interest in basketball sports comparison results P value is less than 0.05, proving that this paper SPOC basketball teaching mode is more conducive to improve the students' basic basketball technical level compared with the traditional teaching mode, and more able to stimulate the students' love for basketball as a sport. The new model of college basketball teaching based on SPOC designed in this paper is of great significance for improving the quality of college basketball teaching.

Index Terms big data analytics, DT-kmeans clustering algorithm, college basketball instruction, SPOC

I. Introduction

The fusion of big data with the realm of education represents a pivotal concept for propelling the profound transformation and advancement of educational systems. This integration stems from the modernization efforts, namely, the reform, innovation, and amalgamation of cutting-edge science and technology, which have diversified the modes of physical education in higher education institutions, thereby offering a multifaceted foundation for enhancing the quality of students' physical education [1-4]. The inherent necessity of this integration between science and technology, and the traditional physical education sector, lies in fostering a comprehensive elevation in the efficiency and quality of physical education. This expansion of physical education modalities aims to uncover more scientific and effective teaching methodologies, emphasizing the distinctive merits of contemporary science and technology in education, and catalyzing the overall reform and development of educational systems [5-8]. Consequently, in the midst of reforming and enhancing basketball teaching in universities and colleges, it is imperative to prioritize adapting to the big data environment and proactively incorporate modern teaching methodologies that align with students' basketball learning needs. Such endeavors will furnish innovative

perspectives for the reform and progression of physical education, ultimately leading to a holistic improvement in students' grasp of basketball's theoretical underpinnings and practical abilities.

The profound implementation of comprehensive quality education has elevated the expectations for physical education and notably enhanced the standing of physical education courses within the academic landscape. Basketball, being a traditional and pivotal discipline within physical education in Chinese universities and colleges, has naturally garnered widespread attention from various stakeholders [9-11]. The teaching of basketball courses in these institutions prioritizes fostering students' foundational basketball competencies. Amidst the advent of the information age and the swift proliferation of big data, the methodologies and modalities of basketball instruction in universities and colleges have undergone significant transformations. From a big data education perspective, the integration of information technology tools with college basketball teaching facilitates the harmonious fusion of basketball talent development objectives, instructional strategies, educational resources, and big data. This integration effectively reorients traditional basketball teaching ideologies and methodologies, thereby strengthening

students' theoretical comprehension of basketball and their proficiency in technical and tactical abilities [12, 13]. Despite the significant advancements in basketball teaching in colleges and universities under the influence and guidance of informatization educational ideologies, an examination of the current state reveals persisting issues such as outdated concepts, inadequate teaching innovations, and a disregard for students' active participation. These issues not only hinder the nurturing of students' interest in basketball but also pose substantial obstacles to the innovative development of basketball education in higher education institutions. In light of this teaching landscape, it is imperative for colleges and universities to devise strategic plans for basketball instruction, meticulously identify shortcomings in basketball teaching amidst the backdrop of big data, and propose innovative strategies for pedagogical enhancement based on an objective analysis of the teaching status quo. Such endeavors will facilitate the substantial development of students' basketball proficiency [14].

Literature [15] employs a multifaceted approach incorporating literature review, questionnaire surveys, and expert interviews to scrutinize the contemporary evaluation framework for basketball teaching outcomes. The study advocates that this system should encompass the attitudes of both students and teachers towards the general basketball course, enhancing instructors' teaching proficiency and students' enthusiasm towards the learning objectives, thereby elevating the quality of students' basketball skills and fostering their physical well-being within the context of a healthily evolving general basketball class. Literature [16] endeavors to delve into the rejuvenation of basketball pedagogical philosophies in higher education institutions. It acknowledges that, despite basketball's technical intricacies, college basketball instruction transcends mere technical proficiency. It emphasizes the cultivation of personality, individuality, and the dissemination of basketball culture through physical conditioning, technical, and tactical training. Furthermore, it highlights the importance of fostering a sense of social responsibility and instilling the virtues of diligence and enterprise. Literature [17] utilizes Visual C++ programming language, OpenGL graphical programming interface, and 3DS Max to implement and refine aspects such as texture mapping, animation, and collision detection across various modules, including the graphical user interface and level design. This technological integration aims to augment athlete training, bolstering their physical conditioning and significantly enhancing their speed and jumping ability in basketball. Literature [18] adopts an ecological perspective, examining the structural and functional dynamics of basketball teaching activities in colleges and universities within the framework of educational ecosystems and ecological balance. It seeks strategies to rectify internal and external ecological imbalances, explores avenues for the sustainable development of basketball teaching in higher education, and offers pertinent recommendations.

Literature [19] introduces an optimization design methodology for a basketball teaching and training system, grounded

in motion capture technology. By juxtaposing the simulation outcomes with authentic athlete training videos, the study demonstrates that basketball motion capture effectively dissects movement discrepancies, facilitating the identification of technical shortcomings and enhancing training efficacy. Literature [20] introduces an optimization design methodology for a basketball teaching and training system, grounded in motion capture technology. By juxtaposing the simulation outcomes with authentic athlete training videos, the study demonstrates that basketball motion capture effectively dissects movement discrepancies, facilitating the identification of technical shortcomings and enhancing training efficacy. Literature [21] discusses the ongoing transformation in the pedagogical philosophy of colleges and universities, highlighting the parallel innovation in the concepts and practices of college basketball programs. It advocates breaking away from conventional teaching models and embracing a health-centric approach to guide the reform of college basketball curricula. The objective is to establish a basketball curriculum model that fosters the holistic development of students' physical and mental wellbeing, addressing the challenges faced by college basketball teaching and enhancing students' overall competencies. Literature [22] examines the practical implementation of sports basketball teaching applications within the contemporary new media landscape. It posits that the utilization of basketball training apps can amplify teaching outcomes, ignite students' enthusiasm for classroom learning, and nurture self-directed learning attitudes. Furthermore, it argues that these apps can augment the effectiveness of traditional basketball training methodologies.

This paper outlines the process of big data analysis, grounded in relevant theoretical frameworks, encompassing data preparation, mining, and subsequent result expression and interpretation. Through an exhaustive examination of the traditional k-means clustering algorithm's underlying principles, we introduce the refined DT-kmeans clustering algorithm. Applying this algorithm, we conduct a clustering analysis of the current state of basketball teaching development in colleges and universities. Building upon the insights gained from this clustering, we incorporate the SPOC teaching model's concept and, leveraging a novel SPOC-based college basketball teaching model, empirically demonstrate its positive impact on enhancing basketball teaching efficacy in academic settings. Furthermore, we delve into the innovative avenues for advancing college basketball teaching methodologies.

II. Data Mining for Basketball Teaching in Higher Education

A. Big data analytics process

Big data analytics encompasses the transformative process of converting vast amounts of raw data into actionable knowledge and insights. This involves a meticulous sequence of steps, including data cleansing, integration, and transformation, to curate datasets suitable for rigorous analysis and min-

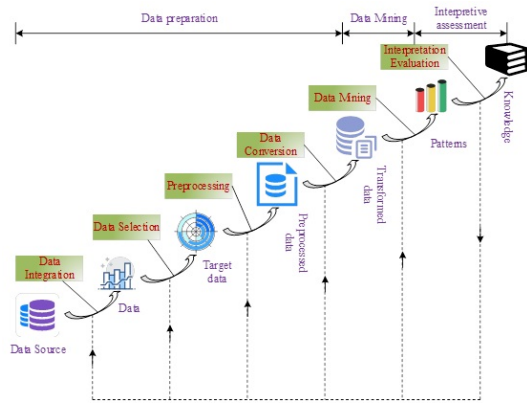


Figure 1: Flowchart of big data analysis

ing. Subsequently, tailored mining models are constructed, leveraging the unique characteristics of each dataset. Through the iterative training and learning process of these models on selected datasets, they are refined to accurately predict, propagate, and correlate pertinent knowledge and data patterns. Ultimately, these insights facilitate informed decision-making. Consequently, the core objective of big data analytics lies in extracting implicit, practical, and efficient knowledge, patterns, and laws from diverse data sources, which are then harnessed to enhance future analytical, predictive, and judgmental capabilities.

The process of big data analytics is shown in Figure 1, which generally includes data preparation, data mining, and result expression and interpretation.

1) Data preparation

Big data analysis is to mine and analyze the data that exists in reality, and in practice, these data are usually used as data stored in the database, which is unclear, missing and contaminated, and usually these data must first be transformed into data suitable for direct mining through certain analytical processing, so the first step of data processing is data preparation. There are three aspects of data preparation specifically, firstly, data integration, secondly, selecting certain data, and finally the detailed pre-processing process. Facts show that data preparation occupies 80% of the workload in the process of big data analysis, and it can be said that it is the focus of the whole big data analysis process and has a very important role. The more detailed the data preparation work is done, the more hassle-free the data mining stage is and the more reliable the conclusions will be.

2) The role of data mining in big data analytics

Data mining constitutes a pivotal stage within the broader knowledge discovery process, marking a critical juncture in the entire big data analytics lifecycle. Notably, it is both the most vital and intricate component of this process, attracting the concentrated efforts of numerous researchers in the field of big data analytics development. During this phase, algorithms such as decision trees, classification and clustering methods,

among others, are employed to delve into processed data, aiming to uncover latent rules and hitherto unknown knowledge patterns.

3) Expression and interpretation of results

Following the rigorous analysis conducted in the preceding stages, we are able to discern and refine valuable knowledge or models from the initial database. Prior to validating the authenticity of these models, it is imperative to assess their merit and relevance. For knowledge deemed congruent with reality and to be of practical value as a model, verifying the consistency between pre- and post-models is crucial to prevent any discrepancies with the original knowledge base. Regarding the newly discovered, applicable knowledge, its utilization must be judicious. Firstly, we leverage the mined knowledge to derive meaningful outcomes, thereby furnishing users with a scientific foundation for decision-making. Secondly, we incorporate the newly acquired data into our model, generating fresh insights and challenges, which we continuously refine and optimize in accordance with predefined objectives. This iterative process repeats until a robust knowledge framework is established.

B. DT-kmeans clustering algorithm

1) Classical k-means clustering algorithm

- a) Principle: The K-means algorithm is a clustering algorithm whose purpose is to divide a set of data into k different clusters, so that the data within the same cluster has a high degree of similarity and the similarity between different clusters is as low as possible. The solution process can be represented by the following steps:
 - Step 1: Randomly select k points as cluster centers.
 - Step 2: For each data point, calculate its distance from the center of each cluster and classify it as the closest family.
 - Step 3: Calculate the center point of each cluster and use it as the new cluster center.
 - Step 4: Repeat steps 2 and 3 until the center of the cluster does not change anymore or a preset number of iterations is reached.

With K-means algorithm, data can be clustered to discover potential patterns and regularities in the data.

- b) Evaluation Indicators: SSE (Sum of Squared Errors): SSE refers to the sum of the squares of the distances of all data points in the clustering results to their corresponding clustering centers. The smaller SSE means the better the clustering effect, which can be used to evaluate the effectiveness of the clustering algorithm. The formula of SSE is:

$$SSE = \sum_{x \in C_i} \|x - \mu_i\|^2, \quad (1)$$

where k is the number of clusters, C_i is the i rd cluster, and μ_i is the center of the i th cluster.

Silhouette Coefficient: Silhouette Coefficient is a kind of index for evaluating the quality of clustering, which can be used to judge whether the clustering effect is reasonable or not. The value of Silhouette Coefficient ranges from $[-1, 1]$, and the closer to 1, the better the clustering effect. The formula of Silhouette Coefficient is:

$$s(x) = \frac{b(x) - a(x)}{\max(a(x), b(x))}, \quad (2)$$

where $a(x)$ is the average distance between data point x and the clustering center it belongs to, and $b(x)$ is the average distance between data point x and its nearest neighbor clustering center.

Calinski-Harabasz Index (CH index): the CH index is a clustering effect evaluation index, which is based on the distance between the SSE of the clustering results and the clustering centers, and the larger means the better the clustering effect. The formula of CH index is:

$$CH = \frac{Tr(B_k)}{Tr(W_k)} \times \frac{n-k}{k-1}, \quad (3)$$

where, B_k is the covariance matrix between the clustering centers, W_k is the covariance matrix between all the data points in the clustering results and the clustering centers they belong to, n is the number of data points, and k is the number of clusters.

Davies-Bouldin Index (DB Index): DB index is a metric used to assess the effectiveness of clustering, which takes into account the distances between clusters and the variability within clusters in the clustering results. A smaller DB index indicates a better clustering effect. The formula for DB index is:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{s_i + s_j}{d(\mu_i, \mu_j)} \quad (4)$$

where S_i is the average distance between the samples of the i th cluster and its cluster center, and $d(\mu_i, \mu_j)$ is the distance between the i th and j th clusters.

2) Improved DT-kmeans algorithm

Given the inherent limitation of the traditional K-means clustering algorithm, which randomly selects initial cluster centroids, leading to instability in cluster division outcomes, this paper introduces an enhanced K-means clustering algorithm aimed at optimizing the selection of initial centroids, designated as the DT-kmeans algorithm. In contrast to the conventional K-means approach, the DT-kmeans algorithm initiates by defining an Eps neighborhood based on the t -nearest neighbor distance. Subsequently, leveraging the statistical density information of data objects within this neighborhood, the algorithm selects an initial centroid from the dataset at random. For the remaining class clusters, centroids are chosen considering the data objects' density and their proximity to the previously determined centroids. Specifically, the selection process prioritizes data objects with high density

and a substantial distance from existing centroids, as determined by a probability function. This ensures that data objects with both characteristics have a heightened likelihood of being selected as new cluster centroids. To facilitate comprehension of the algorithm, the subsequent definitions are presented:

Definition 1. In the Euclidean space R^m , there exist two data objects $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, $x_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}$, and the similarity between them is measured by the Euclidean distance, which is:

$$d(x_i, x_j) = \|x_i - x_j\| = \sqrt{\sum_{i=1}^m (x_i - x_j)^2}. \quad (5)$$

Definition 2. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects and $D_{n \times n}$ is the distance distribution matrix of dataset D , i.e.:

$$D_{n \times n} = \{d(x_i, x_j) | 1 \leq i \leq n, 1 \leq j \leq n\}. \quad (6)$$

Definition 3. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, Eps with the neighborhood parameter η set to:

$$\eta = \lfloor \sqrt[n]{n} \rfloor + 1. \quad (7)$$

Definition 4. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, $D_{n \times n}$ is the distance distribution matrix of dataset D , and the η th smallest distance parameter $d(x_{i\eta})$ is taken out for each row of the distance distribution matrix to obtain the distance array D_η :

$$D_\eta = \{d(x_{i\eta}) | 1 \leq i \leq n, \eta = \lfloor \sqrt[n]{n} \rfloor + 1\}. \quad (8)$$

Definition 5. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, η is the Eps neighborhood parameter and the Eps neighborhood distance is:

$$Eps = \overline{D_\eta} = \frac{1}{n} \sum_{i=1}^n d(x_{i\eta}). \quad (9)$$

Definition 6. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, and the number of data objects in the neighborhood centered on data object x_i and radiused by Eps is called the density $\rho(x_i)$ of data object x_i and is defined as:

$$\rho(x_i) = \sum_{i=1, j \neq i}^n u(Eps - d(x_i, x_j)). \quad (10)$$

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}. \quad (11)$$

Definition 7. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, and the set of class cluster centroids $V = \{v_1, v_2, \dots, v_k\}$ is the set of class cluster centroids for clustering and dividing the dataset, and the Euclidean distance $d(x_i, V)$ between the data objects x_i in the set and the set of class cluster centroids is defined

to be the minimum of the Euclidean distances between x_i and the data objects in the set V :

$$d(x_i, V) = \min(d(x_i, v_j)) \quad j \in (1, k). \quad (12)$$

Definition 8. In the Euclidean space R^m , there exists a dataset $D = \{x_1, x_2, \dots, x_n\}$ containing n data objects, the set of class cluster centroids $V = \{v_1, v_2, \dots, v_l\}$ is the set of data objects that have been selected as class cluster centroids, and the set $T = \{t_1, t_2, \dots, t_{n-l}\}$ is the set of data objects that are not currently selected as class cluster centroids, then the weight $w(t_j)$ of a data object t_j in the set T that has been selected as a class cluster centroid is:

$$w(t_j) = d(t_j, V) * \rho(t_j) \quad j \in (1, n-l). \quad (13)$$

Based on the above definition, we propose a new improved K-means clustering algorithm: the Density K-means algorithm based on the T-th nearest distance. The DT-Kmeans algorithm consists of the following steps in sequence:

- 1) Input the target dataset D containing n data object and the number of class clusters to be clustered in the dataset k .
- 2) Calculate the Euclidean distance between all data objects within the target dataset D according to equation (5), and store the Euclidean distance information in the distance distribution matrix $D_{n \times n}$.
- 3) based on the number of data objects n contained in the dataset, calculate the Eps neighborhood parameter η according to equation (7).
- 4) Based on the distance distribution matrix $D_{n \times n}$ of the data set, take out the η th smallest distance parameter $d(x_{i\eta})$ in each row to get the distance array D_η .
- 5) Based on the distance array D_η , average the distance data in the array according to Eq. (9) to get the neighborhood parameter Eps.
- 6) according to Eq. (10) and Eq. (11), count the density information of the data objects, i.e., the number of data objects in the dataset whose Euclidean distance to the data object is less than or equal to the neighborhood parameter Eps.
- 7) Define an empty set T , and put the data object information in data set D and the density information of the corresponding data object into set T .
- 8) Define an empty set V to hold the class cluster centroids.
- 9) Randomly select a data object from set T and put it into V as the initial cluster center point, and then remove the point from set T .
- 10) Count the minimum value of the Euclidean distance between the data object in set T and the class cluster centroid in set V according to equation (12).
- 11) Select a data object from set T to be added to class cluster centroid set V as a new class cluster centroid. For data object t_i in set T , determine the weight $w(t_j)$ of being selected as the class cluster centroid according to equation (13), and the probability of data object t_i

being added is $p = \frac{w(t_j)}{\sum_{t_u \in T} w(t_u)}$. Remove the data object from set T that was added to class cluster centroid set V .

- 12) Repeating the iterative steps 10 and 11 until the number of data objects in the set V is k .
- 13) Participating in K-means clustering by using the data objects in set V obtained in step 12 as initial class cluster centroids of the K-means clustering algorithm.
- 14) Calculating the distance between each data object in set D and the k class cluster centroids, and assigning the data objects to the class clusters represented by the class cluster centroids with the closest Euclidean distance.
- 15) Count the data object information in each class cluster, take the mean value as the new class cluster center point, and update the class cluster center point information.
- 16) Iteratively perform steps 14, 15 until the algorithm class cluster centroids no longer change.
- 17) Output clustering results: k class clusters independent of each other: $C = \{C_1, C_2, \dots, C_k\}$.

In the execution of DT-Kmeans algorithm, step 1 is the algorithm input and parameter determination stage. Step 2 to step 6 is the data preprocessing stage, where the data object distance information is first calculated and deposited into the distance distribution matrix, and the execution time of this process is roughly $n \times n \times m$, where m is the dimension of the data object. Determine the Eps neighborhood parameter η , the process execution time is roughly: 1. For the distance distribution matrix containing n rows of data, take out the η th smallest distance parameter $d(x_{i\eta})$ in each row to get the distance array D_η , the process execution time is roughly: $n \times n \times \eta$. Determine the neighborhood parameter Eps from the distance array D_η , the process execution time is roughly: n . For the data objects in the dataset, based on the statistical density information of the neighborhood parameter Eps, the execution time of the process is roughly: $n \times n$. Steps 7 to 13 are the initial class cluster centroid selection phase, where the first initial class cluster centroid is randomly selected, and the execution time of the process is roughly: 1, followed by the probabilistic selection of the $k-1$ class cluster centroids, and the execution time of the process is roughly $k \times k \times n$. Steps 14 to 16 are the dataset clustering phase, assuming that the K-means algorithm is executed with an iteration number of 1, the execution time of the process is roughly: $1 \times k \times m \times n$. The computation time of the entire DT-kmeans algorithm is roughly: $n \times n \times (m+n+1) + 1 \times k \times m \times n + n$. The clustering results of the dataset are output in step 17. Typically, all sizes of m, η, k are much smaller than n , so the time complexity of the DT-kmeans algorithm is $O(n^2)$.

The execution of DT-kmeans algorithm is able to respond well to the denseness of the space near the data objects, which reduces the probability of the algorithm falling into the local optimal solution. It improves the clustering quality of the K-means algorithm and the stability of the clustering results.

C. Cluster analysis

	Investigation Item	Completion Criteria
1	Physical Level	0.75
2	Basic Basketball Skill Level	0.80
3	Theoretical Level Of Basketball	0.70
4	Willingness To Learn The Game Of Basketball	0.75
5	Teamwork	0.65
6	Satisfaction With Teaching Methods	0.70

Table 1: Investigate project and completion criteria

1) Data collection on basketball instruction in colleges and universities

The evaluation of basketball teaching results in colleges and universities is of great significance to the quality of physical education teaching in colleges and universities and the improvement of college students' sports literacy. This paper obtains the relevant data of basketball teaching in colleges and universities through a survey questionnaire, the survey object is a college or university to choose the basketball class and the successful completion of the students, the survey questionnaire set up six survey items, respectively, the level of physical fitness, the level of basic basketball technology, basketball theory level, the willingness to learn basketball, the ability to work together as a team and satisfaction with the teaching method. Each item is scored out of ten, and each has a standard value of attainment (0-1), below the standard value of attainment means that the student's ability in this area is weak. The total score is 60 points. Table 1 shows the items and attainment values of this questionnaire.

A total of 350 questionnaires were distributed in the survey and 312 valid questionnaires were retrieved, and finally 300 questionnaire results were retained after data preprocessing as cluster analysis data.

2) Results of DT-kmeans clustering analysis

- DT-kmeans clustering process: Figure 2 depicts the iterative trajectory of the DT-kmeans clustering analysis. Notably, upon the completion of the fourth generation of clustering, the fitness function of the DT-kmeans algorithm attains its initial peak value. However, instead of immediately converging, the algorithm continues to explore, seeking the absolute maximum of the fitness function until approximately the 31st iteration, whereupon it finally converges. This progression exemplifies the salient features of the DT-kmeans clustering algorithm: its rapid convergence and comprehensive global optimization search. Notably, it avoids premature convergence to local extrema, ensuring that the clustering ultimately converges to the global optimum.
- Clustering Results Analysis: Table 2 presents the outcomes of the DT-kmeans algorithm's clustering process, along with the count of distinct clustering samples. The findings indicate that, within the six-dimensional sample space, the DT-kmeans algorithm, after multiple experimental iterations, achieved optimal results with three class centroids. Furthermore, the average values of the six surveyed indicators across these three categories

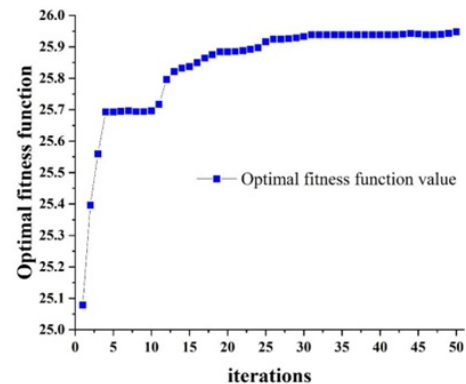


Figure 2: Cluster analysis iteration of DT-kmeans algorithm

	First species	Second species	Third species
Physical level	0.8975	0.7729	0.6933
Basic basketball skill level	0.9036	0.8095	0.7523
Theoretical level of basketball	0.8431	0.7260	0.6508
Willingness to learn the game of basketball	0.9172	0.7642	0.6897
Teamwork	0.7934	0.6823	0.6041
Satisfaction with teaching methods	0.8563	0.7157	0.6539
Sample size	54	214	32

Table 2: Clustering results of DT-kmeans algorithm

were used to categorize basketball course learners into three distinct types, based on the evaluation of teaching effectiveness in basketball classes as measured by these six indicators. As evident from the table, 71.3% of basketball course students fall into the second category, indicating that the overall teaching effect of the current college basketball course is moderate. The first type of students comprises 18% of the total sample and exhibits exceptional physical fitness, advanced basketball skills and theoretical knowledge, a strong proclivity for learning basketball, superior teamwork abilities, and high satisfaction with the current teaching methods. Conversely, the third group constitutes only 10.7% of the sample and displays weak physical fitness, a lack of basic basketball skills and theoretical knowledge, diminished teamwork abilities, low motivation for learning basketball, and dissatisfaction with the current college basketball teaching methods.

III. Innovative Paths for Basketball Teaching in Colleges and Universities

A. SPOC teaching model design ideas

According to the results of the clustering above, the current basketball teaching model cannot be fully satisfied with the students' continued demand. Based on the data, this paper puts forward the strategy of teaching innovation from three stages. The implementation of the SPOC teaching model comprises three distinct phases: the pre-class phase, where students engage in preliminary learning via the internet to gain an understanding of the class content and arrive in class with questions; the in-class phase, wherein students collaborate within the classroom to communicate and resolve issues, thereby

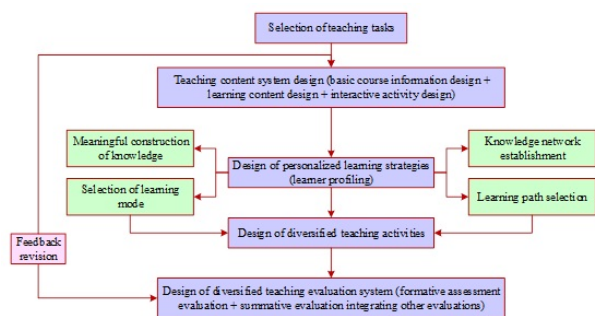


Figure 3: Construction of teaching mode based on SPOC

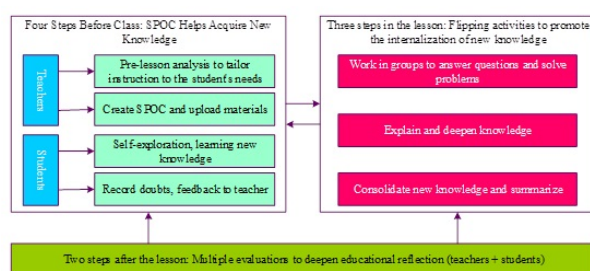


Figure 4: Design drawing of "Four three two" teaching mode based on SPOC

achieving mastery of each knowledge point; and the post-class phase, during which students engage in group learning outside the classroom to facilitate knowledge enhancement. Figure 3 presents the framework diagram outlining the construction of the SPOC-based teaching model, encompassing the design of the teaching content system, learning strategies, teaching activities, and the teaching evaluation system.

The specific idea of this teaching mode is that learners on the SPOC platform can watch short and concise teaching videos from famous schools and teachers, participate in discussions and exchanges, interact effectively with teachers, and get automated feedback after completing assignments on the platform. In addition, the physical classroom gives students a personalized learning experience, followed by review and consolidation, which continuously strengthens their mastery of knowledge. Figure 4 shows the idea diagram of SPOC teaching model, which includes four steps before class, three steps during class and two steps after class, so it is referred to as the "four-three-two model".

B. A new model of college basketball teaching based on SPOC

The traditional way of teaching usually consists of five stages: beginning, preparation, basics, ending and reflection. This section combines the results of the analysis in section 2.3 and still adopts the four steps of teaching in the traditional classroom. Figure 5 shows the step diagram of the new model, which is divided into four steps: preparation for teaching, independent learning before class, practicing and correcting errors during class, and consolidation after class.

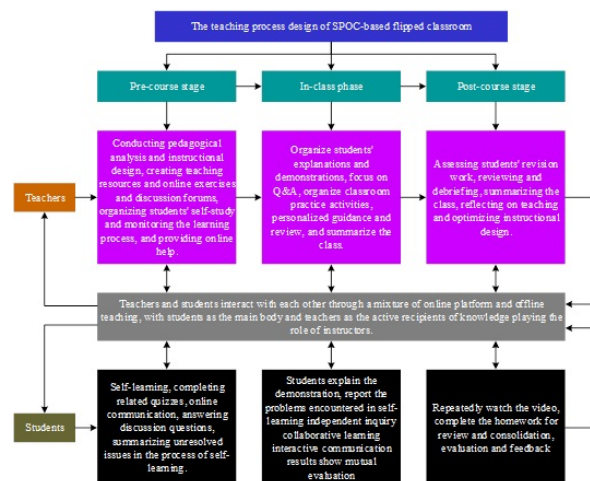


Figure 5: The flow of new teaching mode design

1) Teaching preparation stage

Preparation serves as the cornerstone for implementing the teaching mode. Teachers ought to select suitable learning platforms, establish classes, and, prior to two days before the commencement of the class, develop online teaching resources and assessment methodologies tailored to students' needs, based on the course content, learning objectives, and key points. This enables students to engage in effective learning. Subsequently, the teacher informs students of the course content and requirements by disseminating notifications via the microblogging group. Additionally, students are instructed on how to utilize the learning platform proficiently, ensuring their seamless navigation through the platform.

2) Pre-course independent study stage

In this stage, the teacher uploads the produced video, courseware, PPT, images, and other materials onto the learning platform, where students learn by watching and completing relevant quizzes. The discussion forum is utilized for students to deliberate on problems and engage in mutual communication, thereby fostering their independent learning capabilities. During this phase, students have the flexibility to repeat, adjust, and view the content at a slower pace, allowing for a preliminary comprehension of the technical movements. Simultaneously, the teacher monitors students' learning progress in real-time and promptly addresses issues raised in the discussion area, encouraging peer-to-peer communication to aid in students' knowledge internalization.

3) In-class error correction and practice phase

Contrary to traditional teaching methodologies, teachers allocate a significant portion of time for students to engage in practice, tailored to their individual understanding during the independent learning stage. Teachers emphasize crucial and challenging aspects and subsequently facilitate and organize student practice, incorporating cooperative and group activities to encourage communication and enhance mastery of

technical movements. However, given that students' preliminary understanding of technical movements may be ambiguous, they may superficially mimic movements without fully comprehending their nuances. Consequently, there exists a generalization phase in the learning of technical movements. Teachers must intervene in this process, offering corrections and guidance to assist students in distinguishing and consolidating their skills. During this process, teachers guide students to communicate, identify issues, and showcase exemplary movement practices by select groups, fostering peer evaluation while also offering their own commentary on the group's adherence to technical norms and advantages. Through the meticulous analysis of several incorrect movements, students discern commonalities and identify areas for improvement, thereby deepening their understanding of the skills. Throughout the practice sessions, teachers regularly circulate among students to reinforce their learning.

4) Post-class consolidation stage

The consolidation stage of the class constitutes a pivotal aspect in fostering students' exercise proficiency. The teacher ought to devise practice tasks tailored to the key concepts and difficulties encountered by students in their academic pursuits within the classroom context, and assign a completion timeframe. This enables the teacher to evaluate students' performance accurately. Additionally, the inclusion of technical movement essentials, alongside notes and theoretical knowledge assignments, encourages students to revisit previous videos repeatedly, thereby facilitating the automation of motor skills.

5) Teaching Evaluation

To enhance the scientificity, rationality, and persuasiveness of this teaching mode in terms of students' learning outcomes, a blend of developmental and summative evaluation is employed within the SPOC-based basketball flipped classroom model. Throughout the entire teaching process, which spans pre-class, in-class, post-class activities, and performance assessments, emphasis is placed on leveraging platform evaluations, teacher evaluations, student peer evaluations, and self-assessments. Furthermore, it is imperative to permit experts and teachers to evaluate the effectiveness of the adopted teaching mode, ensuring continuous improvement and optimization.

C. Effectiveness of innovation in basketball teaching in higher education

This section outlines the design and implementation of a 10-week basketball teaching experiment, aimed at assessing the influence of a novel SPOC-based college basketball teaching model on students' learning effectiveness and willingness to learn, both prior to and subsequent to the experiment. The primary objective is to demonstrate the efficacy of this model in enhancing college basketball teaching. The experimental setting is situated within the College of Physical Education at a university in China, with 50 students each assigned

Test project	Experimental group	Control group	Value T	Value P
50m run	6.65±0.37	6.72±0.45	-0.462	0.578
Standing broad jump	2.55±0.14	2.52±0.19	-0.479	0.693
Sit-and-reach	12.37±3.61	12.43±2.98	-0.761	0.726
One minute shot from the free throw line	4.32±1.19	4.41±1.33	-0.296	0.685
Interest in basketball	17.23±2.69	16.89±3.01	-0.467	0.568

Table 3: Test results of experimental subjects before the experiment

to the experimental and control groups. The experimental group will be exposed to the new SPOC-integrated basketball teaching model, whereas the control group will continue with the conventional basketball teaching methods practiced in the institution.

D. Comparison of indicators between the experimental group and the control group before the experiment

To effectively assess the influence of the SPOC-based novel college basketball teaching model on students' teaching experience, both prior to and subsequent to the teaching experiment, we conducted a comparative analysis of the experimental subjects' physical fitness, fundamental basketball skills, and enthusiasm for basketball. This analysis encompassed measurements taken before the implementation of the model, and the specific outcomes of these tests are presented in Table 3.

The data presented in Table 3 demonstrate that, prior to the commencement of the experiment, no statistically significant difference ($P > 0.05$) was observed between the experimental and control groups across three fitness tests: 50m running, standing long jump, and seated forward bending. This signifies that the physical fitness levels of the two groups of students were essentially comparable at baseline. Furthermore, the p-values associated with the pre-experimental comparison of one-minute free throw shooting results and basketball interest between the experimental and control groups were 0.685 and 0.568, respectively. These findings indicate that both groups exhibited comparable levels of fundamental basketball skills and interest in the sport prior to the intervention. In summary, the absence of significant disparities in physical fitness, basic basketball proficiency, and enthusiasm for basketball between the experimental and control groups aligns with the prerequisites for a valid controlled experiment.

E. Comparison of indicators between the experimental group and the control group after the experiment

Upon completion of the 10-week teaching period, the subjects underwent reassessment in terms of their physical fitness, fundamental basketball abilities, and enthusiasm for basketball. The outcomes of these reassessments are presented in Table 4.

Table 4 reveals that the P-values for the experimental and control groups across the three fitness tests—50m running, standing long jump, and seated forward bending—remain above 0.05. Following ten weeks of instruction utilizing the novel SPOC mode, no statistically significant divergence in physical fitness is discernible between the two groups, indicating that, over this brief timeframe, the distinct teaching

Test project	Experimental group	Control group	Value T	Value P
50m run	6.73±0.24	6.74±0.39	-0.621	0.497
Standing broad jump	2.59±0.11	2.54±0.20	0.411	0.618
Sit-and-reach	12.68±3.27	12.49±2.86	0.284	0.649
One minute shot from the free throw line	6.57±1.03	5.05±1.29	2.138	0.026
Interest in basketball	19.46±2.51	17.24±2.99	2.343	<0.001

Table 4: Test results of experimental subjects after the experiment

methodologies have had limited impact on altering students' physical fitness levels. Notably, despite the P-values exceeding 0.05, an intriguing trend emerges when comparing pre-experiment P-values with those post-intervention. Specifically, students in the experimental group have surpassed those in the control group in terms of standing long jump and seated forward bending performance, suggesting that the experimental group's physical fitness may indeed be superior and that the new teaching mode, when applied to basketball classes, holds promise for enhancing students' physical fitness to some extent.

After 10 weeks of teaching experiments, a comparative analysis of the free throw line one-minute shooting test index results between the experimental group and the control group revealed a P-value of 0.026, which is less than 0.05. This statistical significance suggests that there exists a notable difference in the test index outcomes, indicating that the basketball teaching mode based on SPOC employed in this study is more efficacious in enhancing students' basic basketball skill level compared to the traditional teaching mode. Furthermore, a comparative analysis of the interest in basketball, as assessed by the free throw line test, yielded a P-value less than 0.001, substantially below the threshold of 0.05. This marked statistical significance underscores a significant difference in the test index outcomes, suggesting that the SPOC-based basketball teaching mode is more adept at fostering students' passion for the sport of basketball than the conventional teaching approach.

IV. Conclusion

Utilizing the process of big data analysis, this paper employs the DT-kmeans clustering algorithm, an enhancement of the K-means algorithm, to conduct a clustering analysis of the prevailing basketball teaching landscape in colleges and universities. Subsequently, a novel model of college basketball teaching, grounded in the SPOC framework, is devised. The efficacy of this model is rigorously substantiated through a controlled testing regime.

- 1) The clustering algorithm employed categorizes the sampled students into three distinct groups. Notably, 71.3% of the basketball class students belong to the second category as per the cluster analysis results, while 18% of the students fall into the first category, and the remaining 10.7% constitute the third category. These findings indicate that, on an aggregate level, the teaching effectiveness of basketball classes in colleges and universities is positioned at a moderate standard.
- 2) Upon completion of the teaching experiment, a statis-

tical analysis comparing the physical fitness levels of the experimental and control groups yielded a P-value of 0.497, indicating no significant difference between the two groups. In contrast, the P-values obtained from comparing the results of the one-minute shooting test at the free throw line and the interest in basketball between the experimental and control groups were both less than 0.05, signifying a notable difference. These findings suggest that, while the SPOC-based basketball teaching mode employed in this study does not significantly enhance students' physical fitness levels in the short term, it is more effective in elevating their basic basketball skill proficiency and fostering a deeper appreciation for the sport compared to the traditional teaching approach.

References

- [1] Gong, X., & Wang, Y. (2021). Exploring dynamics of sports fan behavior using social media big data-A case study of the 2019 National Basketball Association Finals. *Applied Geography*, 129, 102438.
- [2] González-Espinosa, S., Molina, S. F., García-Rubio, J., Medina, A. A., & García-Santos, D. (2017). Differences in learning according to the teaching method in basketball. *Revista de Psicología del Deporte*, 26(3), 65-70.
- [3] Pauleen, D. J., & Wang, W. Y. (2017). Does big data mean big knowledge? KM perspectives on big data and analytics. *Journal of Knowledge Management*, 21(1), 1-6.
- [4] Gamero, M. G., García-Cebero, J. M., Ibáñez, S. J., & Feu, S. (2021). Analysis of declarative and procedural knowledge according to teaching method and experience in school basketball. *Sustainability*, 13(11), 6012.
- [5] Liu, D. (2018). Introduction to the Training of the Basketball Teaching and Promotion Strategy. *International Journal of New Developments in Engineering and Society*, 2(1).
- [6] Xu, K. (2019, October). The Significance of Injury Prevention Consciousness in Popularizing Basketball Teaching. In 2nd International Conference on Contemporary Education, *Social sciences and ecological studies* (CESSES 2019) (pp. 142-145). Atlantis Press.
- [7] Ge, S. (2017). Research on the factors of basketball injury in physical teaching based on artificial neural network. *Rev. fac. ing.*, 32, 415-422.
- [8] Wang, J. (2019). Training and Reserve of Basketball Talents in China.
- [9] Yu, K., & Feng, K. (2020, August). Research on improving the basketball teaching quality of sports major using modern teaching methods. In *Basic & Clinical Pharmacology & Toxicology* (Vol. 127, pp. 215-215). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.
- [10] Liu, F. (2017). The Application of Hierarchical Teaching Method on Basketball Teaching in Colleges.
- [11] Hai, W. (2021, May). Analysis on the application of network information technology in college public sports basketball teaching. In *Journal of physics: conference series* (Vol. 1915, No. 4, p. 042045). IOP Publishing.
- [12] Silaen, F. B., Kasih, I., & Harahap, N. S. (2021). Development of Macro-media Flash-Based Basketball Teaching Materials for Students' Physical Education, Sports and Health 10th Grade Senior High School Methodist 8 Medan. *Budapest international research and critics institute-journal (BIRCI-Journal)*, 4(2), 2343-2353.
- [13] Arie, S., Ae, M., Fujita, M., Harigai, R., Kashiwagi, Y., & Funato, K. (2019). THE STANDARD MOTION MODEL OF A BASKETBALL SET SHOT FOR TEACHING. *ISBS Proceedings Archive*, 37(1), 77.
- [14] Nugroho, S. A., Soegiyanto, S., & Sugiharto, S. (2019). The Effect of Teaching Models and Motor Skill Levels on The Result of Basketball Learning Technique. *Journal of Physical Education and Sports*, 8(1), 1-6.
- [15] Chen, C., & Zhang, B. (2017). The college basketball teaching evaluation system of fuzzy evaluation. In 2016 *National Convention on Sports Science of China* (p. 01043). EDP Sciences.
- [16] Du, Y., & Liu, B. (2021, May). On the Renewal of Basketball Concepts Based on the Reform of Basketball Teaching in Colleges. In 2021 2nd International Conference on Computers, *Information Processing and Advanced Education* (pp. 959-961).
- [17] Wang, Q., & Wang, X. (2021, January). Application Research of Computer Simulation in Basketball Teaching. In 2021 13th *International conference on measuring technology and mechatronics automation (ICMTMA)* (pp. 788-791). IEEE.

- [18] Kai, H. (2019). Research on the Education Ecology of Basketball Teaching Activities.
- [19] Wang, X. (2017). An optimization method of basketball teaching and training system design based on motion capture technology. *Revista De La Facultad De Ingenieria*, 32(12), 498-504.
- [20] Zhang, N., Han, Y., Crespo, R. G., & Martínez, O. S. (2021). Physical education teaching for saving energy in basketball sports athletics using Hidden Markov and Motion Model. *Computational Intelligence*, 37(3), 1125-1140.
- [21] Liu, S. (2021). Practical research on constructing college basketball curriculum with the concept of health first [J]. *J. Higher Educ. Res*, 2(5), 4410-4423.
- [22] Wei, W., Wei, J., & Peng, J. (2021, February). Practical application of sports basketball teaching app under the background of the new media. In *Journal of Physics: Conference Series* (Vol. 1744, No. 4, p. 042230). IOP Publishing.

...